158841

Gradient Corporation

[redacted]

August 17, 1995

VIA OVERNIGHT DELIVERY

Mr. Jeffrey A. Dodd
U.S. EPA Region III
303 Methodist Building
11th and Chaplin Street
Wheeling, WV 26003

RE: METCOA -- Further Information Regarding 95% UCLM and CRG Statistical Issues

Dear Jeff:

Enclosed is a follow-up discussion regarding the statistical issues which were raised during our recent August 8, 1995 meeting. Specifically, I have addressed the underlying assumptions which are inherent to the calculation of the upper confidence limit on the mean (95% UCLM) and how these assumptions also are linked to the derivation of the confidence removal goal (CRG). The CRG methodology is described in detail in the paper by Bowers et al. (1994), which we provided to you earlier. I would be happy to provide you with any further statistical references if that would be helpful.

I have sent copies of these materials to the individuals indicated below. It is my understanding that Ken Brown will forward a copy of these materials to Dr. Singh.

I look forward to meeting with you so that we can discuss these materials further.

Sincerely yours,
GRADIENT CORPORATION

David E. Merrill

David E. Merrill
Senior Environmental Engineer

enclosure

cc: Catherine Royko
Kathleen Root
Bill Belanger
Ken Brown
Michael Last
Laura Ahern
Rich Kapuscinski

44 Brattle Street, Cambridge, MA 02138 • (617) 57[illegible] • Fax: (617) 864-8469
Cambridge, MA • Ann Arbor, MI • Boulder, CO • Sarasota, FL

METCOA Site
Discussion of Statistical Issues Relative to Action/No Action for Cadmium and Nickel in Surface Soils

During our August 8, 1995 meeting with EPA, a number of questions were raised by EPA, and your statistical expert Dr. Singh of UNLV, which I would like to address. I think it will be helpful to identify those issues where there is general concurrence in the approach, and then address those which remain open with regard to the statistical CRG approach.

Fundamentally, the questions which frame the decision-making on the issue of any further action/no-action for surface soils appear to come down to two very basic ones:

1. ***Is there a need to "clean up?"***
2. ***If so, what concentration values (and locations on the site) must be remediated?***

One could perhaps ask why statistics are needed to help answer these questions. The answer basically is one of practicability and feasibility. Consider, for example, someone who decides to purchase a new car, wanting to find the car at the cheapest price possible. The only unquestionably certain method for finding the "best" price would be to obtain quotes from all possible dealers (nationwide and even worldwide). Unless every possible dealer is contacted, it is impossible to say with absolute certainty that the lowest cost car was purchased. This analogy applies to environmental data. If data were collected from literally every square inch of the site, it would be possible to make decisions with virtual certainty, and leave statistics aside. However, this also is infeasible. Thus, the decision to finally purchase a particular car, or to adopt an action/no-action decision, is based upon gathering sufficient (finite) information to make an informed decision. The assurance that the best decision is reached may be characterized in implicit terms (after calling 10 dealers you were reasonably confident you found the "best" deal), or in explicit terms such as a "statistical confidence level."

In the following discussion, I describe how the statistical analysis of the data provides answers to the two questions raised above. I have attached a diagram which depicts a step-wise decision framework which I believe captures the main elements of the basic decision process as I understood it from our discussions during the meeting. I have also shown that this framework may be based upon either explicit or implicit assumptions, which can have sometimes hidden impacts on the application of the statistics.

It is the specific assumptions which are adopted with respect to the 95% upper confidence limit on the arithmetic mean (95% UCLM), and the assumptions which are embedded in the confidence removal goal (CRG) which I will focus on here.

1. Is there a need to "clean up?"

The answer to this question relies upon evaluating the METCOA Site data relative to risk-based criteria. Because risk-based cleanup goals (TCLs) are determined using the same type approach used in assessing risk, for the purposes of the following discussion it is assumed that meeting TCLs is equivalent to stating the risks for the site are "acceptable" (this would equate to HQ $\leq$ 1 in the case of nickel and cadmium). Furthermore, the discussion assumes the TCL has been established. The TCL itself is subject to judgment (uncertainty) in its derivation, and a further discussion of the appropriate foundations of the TCL with respect to the action/no action decisions will be addressed in conjunction with the statistical issues.

One criteria which EPA indicates it applies to the determination of whether cleanup is required can be summarized as: *Is the 95% UCLM > TCL?*

This decision rule has several key imbedded assumptions.

- *First, it recognizes that the true mean concentration at the Site is uncertain.* Based on a finite sampling of the Site, the sample mean is used to determine the estimate of the true Site mean.

- *Second, a probability distribution must be adopted to quantify the uncertainty in the true mean.* It is not possible to assign a confidence interval on the true mean, without adopting a probability distribution "model."

Because the value of the true Site mean concentration is inherently uncertain, any decision of action or no action will be based on the *probability* of making a correct decision. EPA's guidance indicates that a 95% confidence level is generally the level adopted within the decision framework. In practice, the true confidence level attached to the decision is in the end unknown, because the confidence level depends on how the 95% UCLM is calculated. The appropriate calculation of the 95% UCLM depends on what assumptions are made regarding the distribution of the underlying Site concentration values and the statistical behavior of the sample mean.

Therefore, in order to answer the question of whether cleanup is required, one or more assumptions must be made regarding the calculation of the 95% UCLM. As discussed at the August 8, 1995 meeting, I understood there to be general agreement regarding the following:

- EPA guidance[1] provides methods for calculating a 95% UCLM using the "t" and "H" statistics.[2]

---

[1]U.S. EPA. May, 1992. *Supplemental Guidance to RAGS: Calculating the Concentration Term.*

[2]It is important to remember that the 95% UCLM is the upper confidence limit on the arithmetic mean. Statistical procedures are available to determine confidence limits on the distribution parameters for other than the normal and lognormal distributions, however they are not relevant to the arithmetic mean.

- The W-test and K-S test used to assess the fit of the data to normal and lognormal distributions indicate cadmium passes both tests for lognormality, and nickel passes the K-S test for lognormality (EPA indicated nickel also passed the D'Agostino test).

- The sample mean concentration for cadmium and nickel for the combined fenced and perimeter area is lower than both the EPA TCL and the MO/AR TCL

- The 95% UCLM using the t-statistic is below the MO/AR TCLs for both cadmium and nickel, and also below EPA's TCL for nickel.

- The 95%UCLM using the H-statistic is above the TCLs for both cadmium and nickel.

The foregoing results indicate that the answer to the question of whether there is a need for clean up depends upon the assumptions regarding the data as summarized below:

Statistic	*Conclusion*	*Assumption*
Mean < TCL	no action .	No distribution assumption
95% UCLM < TCLs* (t-statistic as in the MO/AR)	no action	Site concentrations follow any distribution; Sample mean is approximately normal.
95% UCLM > TCLs (H-statistic)	Possibility that true mean exceeds TCL	Site concentration values are lognormal
*only cadmium exceeds the EPA TCL for this scenario.		

Thus, a "yes" answer to the first question, "is clean up required" is based upon adopting the lognormal distribution to fit the data:

> Only if the lognormal assumption is adopted is the conclusion reached that the H-statistic 95% UCLM > TCL and there is a possibility the true mean exceeds the TCL. This possibility cannot be eliminated with 95% certainty based on this criterion.

As will be discussed later, the H-statistic 95% UCLM is as large as it is due to the large skew in the data. To the extent that the lognormal distribution assumption overpredicts the actual skew in the Site data (it assigns greater probability to values which go to infinity than is physically correct) it leads to an overprediction of the 95% UCLM. The impact of assuming the Site concentration data are lognormal and hence using the H-statistic to calculate the 95% UCLM is readily apparent from the comparison of the 95% UCLM values calculated using the t-statistic and H-statistic:

Chemical	Mean	t-statistic 95% UCLM	H-statistic 95% UCLM
Cadmium	613.7	923.4	1,733 (92% < UCLM)
Nickel	5,341.0	7,493.4	33,458 (96% < UCLM)

As noted above, 92% and 96% of the actually measured cadmium and nickel sample concentration values, respectively, fall below the H-statistic 95%UCLM.

Why is the H-statistic 95% UCLM so far out on the tail of the distribution? Essentially, the reason is that the lognormal assumption implies values within the distribution that go to infinity. Thus, values exceeding 1,000,000 mg/kg (pure compound) are assumed to occur within the fitted distribution. ***This is clearly an artificial assumption, with no real-world meaning***. It is this assumption that forces the 95% UCLM out on the tail of this lognormal distribution. For example, the maximum observed sample concentration for nickel (69,800 mg/kg) has a probability of 2.1% of being exceeded based on the lognormal distribution used to calculate the UCLM. Yet, the empirical probability (simply based on the observed data) of this value being exceeded is approximately 1%. The concentration corresponding to the 1% exceedance value for the "fitted" lognormal distribution is 147,573 mg/kg. Thus, the fitted lognormal distribution predicts more than a two-fold greater chance of observing values exceeding the maximum observed value than if the Site data alone are used.

What this says is that the lognormal "model" fit to the sample data imposes a greater skew to the large values than is observed in the actual Site data. This skew forces the UCLM to a value which is greater than it would be if a distribution were chosen which reflects that concentration values cannot be greater than 100%.[3] While it is true that the goodness of fit tests indicated the lognormal distribution provided an adequate fit to the data, it is also true that this assumption has the inherent limitations just described. Several approaches offer possible alternatives to help apply some "ground truthing" to the above issue:

1) fit a "better" distribution which does not assume values that go to infinity and recalculate the 95% UCLM for this distribution,

2) use the H-statistic 95% UCLM together with the CRG, understanding that the H-statistic UCLM overestimates the confidence interval on the mean, or

3) adopt the t-statistic 95% UCLM, as was done in the MO/AR, which assumes the mean is approximately normal (not the data themselves).

[3]In reality the absolute upper limit for soil matrices would be less than 100%, because at 100% chemical concentration there would no longer be a soil matrix. A more likely physical upper bound would be on the order of 30% - 50% chemical concentration (300,000 - 500,000 mg/kg).

While the first alternative has broad-based application, it may not be feasible within the scope of this project. Currently, EPA guidance only provides for using the t-statistic or H-statistic when calculating the upper confidence limit on the arithmetic mean. If the second alternative is adopted (e.g., the H-statistic 95% UCLM is used), then the assumption of a lognormal distribution of the Site data applies and it follows that the CRG approach defines the concentration of chemical which must be targeted for remediation. Because the same lognormal distribution assumptions underlie the H-statistic UCLM and the CRG calculation, the two are consistent and inescapable conclusions resulting from the lognormal model.

A third alternative, would be to adopt the use of the t-statistic established in the MO/AR. As stated earlier, the use of the t-statistic rests on the assumption that the sample mean is approximately normal, a fact that follows from the Central Limit Theorem which states:[4]

> Central Limit Theorem: For sufficiently large numbers of samples (N), the mean value of a random variable has an approximately normal distribution.

It is generally considered, that for N> 30, the Central Limit Theorem can be used. The assumption of normality of the mean holds, even when the underlying distribution is highly nonnormal, thus *requiring no specific assumption about the distribution of the sample data*, other than that they are statistically independent (Devore, 1987).

The key distinction between the above assumption and the earlier discussion of lognormality of the Site data set, is that the Central Limit Theorem applies to the sample mean, not to the distribution of the underlying sample values. Therefore, the Central Limit Theorem provides an alternate avenue for determining the upper confidence limit on the true Site mean. Adopting this approach, the 95% UCLM is calculated using the t-statistic as was done in the MO/AR and as we have summarized in earlier meetings. There are 97 and 98 samples respectively in the cadmium and nickel data sets. This sample size is well above the 30 samples indicated above as providing a generally sufficient number to adopt the assumption that the sample mean is normally distributed and hence that the t-statistic 95% UCLM is valid.

Summary of Question #1:

- Assuming the Site concentration values are lognormaly distributed, the H-statistic used to calculate the 95% UCLM is greater than the TCLs.
- The H-statistic 95% UCLM is biased high due to the underlying assumptions which break down at the far ends of the tail of the distribution for real environmental data.

[4]Devore, J.L., 1987. Probability and Statistics for Engineering and the Sciences, 2nd ed. Brooks/Cole Publishing Co., Monterey CA.

- The assumption of lognormal Site concentration values underlies both the calculation of the H-statistic 95% UCLM and the calculation of the CRG. One cannot accept the results of one while disregarding the other.

- Use of the t-statistic to calculate the 95% UCLM is justified. The 95% UCLM under this assumption indicates no action would be necessary for nickel for both the EPA and MO/AR TCLS; no action would be required for cadmium under the MO/AR TCL.

2. *What Concentration (and Areas) Would Require Remediation if Cleanup is Required?*

As discussed during our meeting, if remediation is required, the level of remediation is targeted to achieving the TCLs on average. Because the true average (mean) is uncertain, the true mean is conservatively assumed to be as high as the 95% UCLM.[5] Thus, the goal of remediation is to achieve an estimate of the true mean value which is below the TCL with 95% confidence. As will be explained below, this goal is achieved when values above the CRG are "removed" from the lognormal distribution.

Achieving the target cleanup level does not require all values above the TCL to be removed, only those "high" values which cause the estimate of the true mean to exceed the TCL with at most a 5% probability. The CRG approach was developed to quantify the threshold that defines the "high" concentration values in relationship to the 95% UCLM and the TCL. Without such an approach, there is no clear mechanism to determine what concentration is "high" enough to cause the estimate of the true mean to exceed the TCL.

The important connection between the H-statistic 95% UCLM and the CRG is that both rely upon the assumption that the underlying concentration values are lognormal. The CRG defines the point on the lognormal distribution such that all values above the CRG must be removed (truncated) in order to reduce the estimate of the true mean to below the TCL (with 95% probability). The derivation of the CRG has been peer reviewed, and the paper describing this derivation has been accepted for publication in *Environmental Science and Technology*.[6] As described on pages 11 and 12 of the paper by Bowers *et al.* (1994), the CRG approach is the statistically correct means to determine a "not to be exceeded value" when decisions are based on the UCLM, as opposed to the mean.

---

[5]U.S. EPA, May 1992, *op.cit.*

[6]Bowers, T.S., N.S. Shifrin and B.L. Murphy. 1994. Applying hazardous waste site cleanup levels: A statistical approach to meeting soil cleanup goals on average. *Environmental Science & Technology* (accepted for publication).

If the 95% UCLM > TCL, isn't cleanup required?

Not necessarily. The fact that the 95% UCLM is above the TCL only indicates there is a ***possibility*** that the true mean exceeds the TCL, and some remediation may be required. The lognormal distribution fit to the data, which is the basis for concluding the UCLM exceeds the TCL, presumes there are concentration values extending to infinity. ***When concentration values (fitted to the lognormal model) exceeding the CRG do not exist (or are remediated), there is a 95% probability that the true mean is below the TCL.*** Therefore, only those concentration values above the CRG require cleanup to achieve the TCL with 95% confidence.

It may be tempting to disregard the CRG method and fall back on the sample data to determine the "not to be exceeded" value, or to recalculate the 95% UCLM in an iterative manner. Yet, as indicated in Bowers *et al.* (p.11 - 12), a brute force recalculation of the mean and 95% UCLM by successively removing the highest values isn't strictly valid. By definition such censorship of the data will violate the random and independent assumptions used to calculate the 95% UCLM, and also will violate the assumption that values exist higher than those sampled -- indeed the lognormal distribution assumes values that extend to infinite concentration.

Discussion of Kriging Alternative

During the August 8, 1995 meeting with EPA, Dr. Singh mentioned that kriging offered an option to define areas of "high" concentrations, and is a standard method which could be used to determine areas for cleanup. It is true that kriging is a powerful geostatistical tool that can be used to help define the distribution of chemical concentration at the Site. Kriging is essentially a sophisticated method of interpolating chemical concentration contours, providing a weighted concentration contour map. The interpolation is "weighted" in the sense that concentration is assumed to be correlated in space, where the correlation typically is greater for samples spaced closely and declines with distance for samples that are widely spaced.

As powerful as this tool is, it still does not provide the answer to the second question raised in this memorandum: How high is "high?" In other words, kriging will show the areas of high and low concentration at the Site, but it comes no closer to defining which high values must be remediated such that the 95% UCLM ≤ TCL. Without further development of the kriging method, it alone does not provide the information which would yield the "not to be exceeded concentration" in order to meet the TCL. Kriging could be used to refine the calculation of the 95% UCLM (*e.g.*, providing an alternative assumption to those described in this memo). However, this would require developing the method further to determine the variance of the kriged concentration estimates in order to resolve the bias in the existing sampling.

3. *Summary*

A decision regarding further action/no action for surface soils relies upon several complex statistical issues leading to the determination of the 95% UCLM and comparing this to the TCL. Both the UCLM and the TCL have inherent assumptions which must be examined in order to determine the technical basis for a decision regarding surface soils at the METCOA Site. This memorandum has addressed the statistical issues involved in the decision process (issues regarding establishing a site-specific TCL are addressed in a separate document) and establishes the following points:

- Nickel and cadmium in surface soils would not require further action if the t-statistic 95% UCLM is adopted and the MO/AR cleanup levels are adopted; nickel would not require further action for either the EPA or MO/AR TCLs.
- The t-statistic 95% UCLM was adopted in the MO/AR, and it is valid regardless of the underlying distribution of the data because the sample size is large.
- The H-statistic 95% UCLM is biased high, and any decision based on this statistic alone is biased high.
- The H-statistic 95% UCLM and CRG calculations are fundamentally based on similar assumptions; adopting the H-statistic 95% UCLM as "correct," leads inescapably to adopting the CRG value also.
- None of the cadmium and nickel sampling data exceed the calculated CRG values, indicating that none of the currently identified surface soil areas require further action.

As summarized above, the "no further action" alternative for nickel and cadmium in surface soils is reached no matter whether the t-statistic or the H-statistic assumptions are adopted. If the t-statistic assumptions are adopted, no "new" methodology needs to be adopted, and this approach is consistent with the analysis in the MO/AR. The CRG methodology has been peer reviewed and provides a powerful tool to determine appropriate action levels which are risk-based and cost-effective.

Both of these approaches are valid; they simply differ in their underlying assumptions. As long as clear decision rules are followed, either of them can be applied so long as each is applied consistently.

Why are statistics required?	Site characterization is based on a finite data set and the certainty, and uncertainty, of decisions can only be quantified explicitly using statistical methods.
What statistical models of the uncertainty in the mean are supported by EPA?	EPA describes methods of calculating the 95% UCLM using both the t-statistic and the H-statistic.
Which distribution should be used?	A lognormality test can be performed. EPA indicates that a lognormal assumption typically applies to environmental data.
Can the t-statistic UCLM be used regardless of the underlying distribution of the data?	Yes. The *Central Limit Theorem* in statistics indicates that the mean of a random variable tends toward a normally distributed variable, as the number of samples increases. This holds even when the underlying data set is skewed. Typically, when N>30 samples, a normal approximation for the mean can be used; there are >95 samples for the METCOA data set.
Why does the 95% UCLM for a lognormal distribution sometimes yield results which are counter-intuitive (*e.g.*, the 95%UCLM for nickel is approximately the 95th percentile of the measured data)?	The lognormal distribution assumes a "skewed" data set, with values starting at zero (0) and going to infinity. Because the lognormal distribution assigns a relatively high probability to large values, the H-statistic 95%UCLM can rapidly become extremely large, in some cases even larger than the maximum value detected (as recognized in EPA guidance).
What does it mean when the mean is below the TCL, yet the 95% UCLM is above the TCL?	This result indicates that there is some possibility the true mean is above the TCL and thus some possibility that some action is required, there is also the possibility the true mean is below the TCL and no action would be required.
When is the TCL "achieved?"	The TCL must be met on average. Due to uncertainty in the average (mean), the TCL is considered achieved when the 95% UCLM ≤ TCL. ***Meeting the TCL does not require removing all values above the TCL.***
What is the CRG and how is it related to the UCLM?	The CRG is calculated based on the same assumption underlying the H-statistic UCLM: namely the site concentration data are lognormal. The CRG is the point on the lognormal distribution that will yield a 95% UCLM ≤ TCL if all values above the CRG are "removed" from the distribution (this truncates the distribution) and replaced with values with nominal concentration.
The H-statistic 95%UCLM is above the TCL, doesn't this mean soil must be remediated?	Not necessarily. Only values above the CRG require remediation in order to meet the TCL. The lognormal distribution fit to the data, which is the basis for concluding the 95% UCLM > TCL, presumes there are concentration values extending to infinity. It is only the actual values above the CRG which would cause the true mean to exceed the TCL; only those > CRG must be remediated.